Linguistic Bayesian Networks for Reasoning with Subjective Probabilities in Forensic Statistics

Joe Halliwell
Centre for Intelligent Systems and their Applications
The University of Edinburgh
Appleton Tower, Crichton Street
Edinburgh, EH8 9LE, UK
J.L.Halliwell@sms.ed.ac.uk

Jeroen Keppens
Centre for Forensic Statistics and Legal Reasoning
The University of Edinburgh
Appleton Tower, Crichton Street
Edinburgh, EH8 9LE, UK
jeroen@inf.ed.ac.uk

Qiang Shen
Centre for Intelligent Systems and their Applications
The University of Edinburgh
Appleton Tower, Crichton Street
Edinburgh, EH8 9LE, UK
qiangs@inf.ed.ac.uk

ABSTRACT

Recent work in forensic statistics has shown how Bayesian Networks (BNs) can be used to infer the probability of defense and prosecution statements based on forensic evidence. This is an important development as it helps to quantify the meaning of forensic expert testimony during court proceedings, for example, that there is "strong support" for the defense or prosecution position. Due to the lack of experimental data, inferred probabilities often rely on subjective probabilities provided by experts. Because these are based on informed guesses, it is very difficult to express them accurately with precise numbers. Yet, conventional BNs can only employ probabilities expressed as real numbers. To address this issue, this paper presents a novel extension of probability theory. This allow the expression of subjective probabilities as fuzzy numbers, which more faithfully reflect expert opinion. By means of practical examples, it will be shown that the accurate representation of this lack of precision in reasoning with subjective probabilities has important implications for the overall result.

Keywords

Linguistic Probabilities, Decision Support Systems, Crime Investigation

1. INTRODUCTION

The question of how to avoid miscarriages of justice has concerned governments and justice systems for many centuries. With the ever increasing public and media scrutiny and the continued introduction of new types of forensic evidence, dealing with this issue has not become any easier.

Forensic statistics has emerged as an important discipline in this respect by providing techniques, such as the likelihood ratio [2], to evaluate evidence in terms of its relative support of claims made by the prosecution vs. claims made by the defence. In other words, these methods provide a statistical foundation for the expert testimony of expert witnesses. (??)

A central component of this work is the use of Bayesian Networks (BNs) to compute the probability $P(E \mid C)$ of obtaining certain pieces of evidence E given a claim C. An example of $P(E \mid C)$ is the probability of finding a certain number of glass fragments in the clothes of a person assuming that that person has smashed a window. This is not a trivial task because many factors influence the production of evidence. In the glass fragment example, the number of glass fragments retrieved in the laboratory depends on the way the window was smashed, the movements of the perpetrator after the crime (which cause some of the glass fragments to fall from the garment) and the laboratory techniques employed. BNs provide an effective way of organising this knowledge. Additionally, they also enable the use of efficient algorithms to compute the probability of interest.

Common criticisms of the Bayesian approach are that it requires too many numeric probability estimates and the probabilities are often subjective estimates by experts.

The difficulty of obtaining point estimates of (e.g. prior) probabilities in general has been widely reported ([13, 18]). Moreover [19] has reported that verbal expressions of probabilistic uncertainty were more accurate than numerical values in estimating the frequency of multiple attributes in his experimental studies. In addition subjective probability assessments are not generally precise and it has been claimed that it is misleading to seek to represent them precisely. So, for example, a committee of the U.S National Research Council has written that there is an

> an important responsibility not to use numbers, which convey the impression of precision, when the understanding of relationships is indeed less secure. Thus whilst quantitative risk assessment facilitates comparison, such comparison may be illusory or misleading if the use of precise numbers is unjustified.[1]

All this suggests that it would be useful to involve probabilistic terms directly in probabilistic models. Various studies (e.g. [3]), have concluded that point estimates of probabil-

[1] [14] as quoted in [16].

ity terms are highly variable inter-subjectively and exhibit great overlap between terms.

This paper presents a novel approach to the representation of subjective probability assessments known as linguistic probabilities. Fuzzy sets have been widely used to represent the inherent vagueness in linguistic descriptions. Furthermore, a number of psychometric studies have evaluated the claim that fuzzy sets may be used to model qualitative probabilites with generally positive conclusions. So, for example, [16] have considered various methodological issues in detail, and established that experimentally obtained fuzzy sets do indeed seem to provide a model for every day probabilistic assessments.

Linguistic Probability Theory [11] combines the ability of fuzzy sets to capture the haziness of natural language, with a calculus modelled after that of classical probaility theory. The result is a formal setting for computing with probabilities which may be defined by qualitative expressions such as "likely", or "nearly certain". As an extension of classical probabilities, linguistic probabilities can represent everything from a complete lack of knowledge about the chance of some event's ocurring (every probability is judged equally appropriate) to precise (perhaps experimentally determined) knowledge (e.g. the probability of event X is 0.43452561).

Specifying probabilities qualitatively can facilitate the notoriously time-consuming and error-prone knowledge acquistion process. Moreover, whilst the experts are likely to be statistically sophisticated enough to understand numbers, the classical approach does not allow them to express any uncertainty they may have[2]. In other words, experts are usually capable of producing accurate probability estimates, but not precise ones. Therefore, the issue of using BNs with subjective probabilities is not due to doubts about the accuracy of the results, but a concern for the false impression of precision that this approach implies.

Finally, the greater interpretability of linguistic probabilities carries through to the results of computations. This has obvious promises in a legal context where derived probabilistic information must be presented in a form that is intelligible and persuasive.

The remainder of the paper is organised as follows. The following section introduces the application of Bayesian inference in forensic statistics and presents a detailed and realistic example scenario which will be used throughout to provide a common basis for evaluating the proposed extension of classical probability theory. [It closes with motivation.] Section Two begins with a brief introduction to fuzzy set theory and recapitulation of classical probability theory before showing how these may be combined to yield LPT. Section Three develops the two applications and exhibits the related sample computations. Section Four concludes the paper and indicates areas for future work.

2. THE ROLE OF BAYESIAN INFERENCE IN FORENSIC STATISTICS

Bayesian Networks (BNs) are an effective tool in forensic statistics to help determine to what extent forensic evidence support the claims of the prosecution (defense) vs. those of the defense (prosecution). This section describes how BNs are used to that end and illustrates how they are constructed by means of a representative example (QS: Make clear that we're not talking about KA or Learning.). Then, the difficulty of obtaining point estimates of probabilities in such BNs will be discussed to demonstrate the need for linguistic BNs. Note that a detailed description of BNs is beyond the scope of this paper and the reader is referred to [15] for a more detailed discussion.

[2]A second-order Bayesian approach would be *better* equipped to handle such uncertainties, but is semantically problematic and typically computationally intractable.

LR	Support of evidence to prosecution claim over defense claim
1 to 10	limited
10 to 100	moderate
100 to 1,000	moderately strong
1,000 to 10,000	strong
$> 10{,}000$	very strong

Table 1: Interpretation of the likelihood ratio

2.1 Background

Forensic statistics is a discipline that is mainly concerned with the experimental design of forensic examinations and the analysis of the obtained results. The issues it studies include hypothesis formulation, deciding on minimal sample sizes when studying populations of similar units of evidence and determining the statistical significance of the outcome of tests. Recently, the discipline has been branching out to the study of the statistical implications of forensic examinations on defense and prosecution positions during crime investigation and criminal court proceedings.

In [5], a method is proposed to assess the impact of a certain piece of forensic evidence on a given case. This method is the result of a significant research effort by the Forensic Science Service (FSS), the largest provider of forensic science services in England and Wales. It involves 1) formalising the respective claims of the prosecution and the defense [4, 8], 2) computing the probability that the evidence is found given that the claim of the prosecution is true and the probability that the evidence is found given that the claim of the defense is true, and 3) dividing the former probability by the latter to determine the likelihood ratio [2]:

$$LR = \frac{P(E \mid C_p)}{P(E \mid C_d)}$$

where LR is the likelihood ratio, E, C_p, C_d respectively represent the evidence, the prosecution claim and the defense claim, and $P(E \mid C)$ is the probability that evidence E is found if claim C is true. The likelihood ratio is a numerical evaluation of the extent to which the evidence supports the prosecution claim over the defense claim. It has two important application. Firstly, expected improvement of the likelihood ratio due to a, potentially expensive and resource intensive, forensic examination represents the benefits of this examination. As such, the likelihood ratio can help the decision making by police forces when considering purchasing forensic services. Secondly, the likelihood ratio can be used to justify the testimonies of forensic experts during the court proceedings. To that end, a verbal scale to help forensic experts interprete LR is suggested by the FSS in [9] and reproduced in table 1.

The likelihood ratio method is, of course, crucially dependent upon a means to compute the probabilities $P(E \mid C_p)$

	Event	Domain
q_t	quantity of transferred fragments	{none,few,many}
q_p	quantity of persisted fragments	{none,few,many}
q_l	quantity of lifted fragments	{none,few,many}
t_c	type of contact	{none,some}
p_s	proportion of shedded fragments	{none,small,large}
p_l	proportion of lifted fragments	{some,most,all}

Table 2: Variables in the one-way transfer case

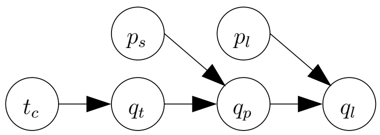

Figure 1: Bayesian network of a one-way transfer case

and $P(E \mid C_d)$. As shown in [1, 6, 7], Bayesian Networks (BNs) are particularly suitable determine such probabilities. They are an effective means of acquiring and representing the knowledge required to compute these probabilities and they enable the use of efficient algorithms to calculate them.

A BN is directed graph in which events are denoted as nodes and causal relations are denoted as arcs from cause to effect. The lack of a path from one node to another represents conditional independence. This enables BNs to be used as a tool for knowledge acquisition and probability calculations. An example may best illustrate the use of Bayesian networks in this context.

Consider the following scenario.

> A burglar smashes the window of a shop, steals some money from the cash registry and flees the scene of the crime. A bystander witnessed this event and reports a description of the perpetrator to the police who arrest a man, matching the description of the witness half an hour after the event. The suspect, Mr. Blue, denies having been near the shop. However, q_l glass fragments, matching the type of glass of the shop's window, are retrieved from Mr. Blue's clothes.

Let E be the retrieval of q_l glass fragments from the clothes of Mr. Blue and let C_p be the assumption that Mr. Blue was the person who smashed the window of the shop.

In this case, figure 1 shows a BN for computing the probability $P(E \mid C_p)$, where E is the retrieval of q_l glass fragments from the garment of Mr. Blue in the forensic laboratory and C_p is a presumed type of contact between Mr. Blue and the shop's window. This BN represents the following forensic knowledge. The number of glass fragments q_l that are retrieved from Mr. Blue's clothes depends on the number of glass fragments that have persisted in the clothes q_p and on the effectiveness of the retrieval technique p_l, where p_l represents the proportion of glass fragments lifted from the garments under examination. The number of glass fragments q_p that have persisted in the clothes until the time of the exampination, in turn, is dependent upon the number of glass fragments q_t that were transferred in the first place and the proportion of fragments p_s that were shedded between the time of transfer and the time of the examination. Finally, the number of transferred fragments

p_s	$\mathrm{P}(p_s)$
none	0.1
small	0.4
large	0.5

t_c	$\mathrm{P}(t_c)$
none	0.6
some	0.4

p_l	$\mathrm{P}(p_l)$
none	0.1
few	0.3
many	0.6

Table 3: Classical prior probabilities for $\mathrm{P}(p_s), \mathrm{P}(t_c)$ **and** $\mathrm{P}(p_l)$**.**

q_t depends on the type of contact t_c. The BN of figure 1 can now be described mathematically by the following equations:

$$P(q_t) = \sum_{t_c} P(q_t \mid t_c) P(t_c)$$

$$P(q_p) = \sum_{q_t} \sum_{p_s} P(q_p \mid q_t, p_s) P(q_t) P(p_s)$$

$$P(q_l) = \sum_{q_p} \sum_{p_l} P(q_l \mid q_p, p_l) P(q_p) P(p_l)$$

For given values of t_c, BNs can collapse the computation the conditional probability $P(q_l \mid t_c)$ of retrieving q_l fibres from the clothes of Mr. Blue given a type of contact t_c to:

$$P(q_l \mid t_c) = \frac{P(t_c) \sum_{p_l} P(p_l) \sum_{q_p} P(q_l \mid q_p, p_l) \sum_{q_t} P(q_t \mid t_c) \sum_{p_s} P(q_t \mid q_p, p_s) P(p_s)}{P(t_c)} \quad (1)$$

Because $P(q_l \mid t_c)$ equals $P(E \mid C_p)$, the latter equation enables the computation of the required information of the probability of obtaining a given piece of evidence assuming that the prosecution claim is true.

2.2 Motivation

It follows from (1) that the calculation of $P(q_l \mid t_c)$ requires numeric estimates for $P(p_l)$, $P(q_l \mid q_p, p_l)$, $P(t_c)$, $P(q_t \mid t_c)$, $P(q_t \mid q_p, p_s)$, and $P(p_s)$ for all possible values of p_l, p_s, q_l, q_p, q_t and t_c. Such values can be obtained through experimentation. For example, $P(q_t \mid t_c)$ can be determined by smashing a representative population of glass panels of the same make and materials as the shop window with a piece of fabric similar to that of the garment worn by the suspect. Then, the distribution of the number of glass fragments transferred to the pieces of fabric throughout the experiment may provide a reasonable estimate for $P(q_t \mid t_c)$.

Such experiments are obviously very expensive. Therefore, it is often necessary to rely on estimates provided by experts.

Suppose that the prosecution case is that the defendent has had *full* or *partial* contact with the window in question and that a forensic procedure which is known to retrieve *most* fragments of glass from a garment has yielded a *few* matching fragments.

The probabilies required to evaluate the likelihood ratio are provided in Tables 4, 5 and 6. Note that the prior distribution of p_l is not required as these terms cancel in the calculation.

3. LINGUISTIC PROBABILITY THEORY

There have been a number of attempts to provide a theory of fuzzy probabilities ([17], [12]). For technical reasons,

t_c	P($q_t = none \mid t_c$)	P($q_t = few \mid t_c$)	P($q_t = many \mid t_c$)
light	0.2	0.6	0.2
medium	0.1	0.4	0.5
heavy	0.02	0.28	0.7

Table 4: Classical conditional probabilities of $P(q_t \mid t_c)$.

q_t	p_s	P($q_p = none \mid q_t, p_s$)	P($q_p = few \mid q_t, p_s$)	P($q_p = many \mid q_t, p_s$)
none	none	1	0	0
	small	1	0	0
	large	1	0	0
few	none	0.01	0.99	0
	small	0.2	0.8	0
	large	0.5	0.5	0
many	none	0	0.01	0.99
	small	0.01	0.14	0.85
	large	0.1	0.45	0.45

Table 5: Classical conditional probabilities of $P(q_p \mid q_t, p_s)$.

however, these exististing formalisms are incapable of expressing the fuzzy models for qualitative probabilities obtained in the psychometric literature cited in Section 1[3] . Linguistic Probability Theory represents a new approach to a qualitive probability theory centered on fuzzy numbers which overcomes these limitations.

A considerable amount of background material is required to explain and motivate Linguistic Probability Theory. While it is assumed that the reader will have a passing familiarity with at least some of this material, the basic concepts and definitions are rehearsed below. In the following section fuzzy sets and logic are introduced as a prelude to the formal definition of fuzzy numbers and their associated arithmetic operators and partial orderings. Section 3.3 then briefly recapitulates the entities and axioms of classical probability theory. With these preliminaries in place the final section presents Linguistic Probability Theory itself.

3.1 A brief introduction to fuzzy sets, logic and numbers

Fuzzy sets were introduced as a means of capturing the vagueness of everyday concepts. Consider the concept of redness. Certain objects, such as postboxes and strawberries are indisputably red. Conversely, there are things such as this sheet of paper that are certainly not red[4]. There are however many things that are merely somewhat red: claret, the human tongue and auburn tresses. In these cases, classical logic with its binary concept of set membership[5] forces a kind of universal judgement call, requiring us to decide yes or no whether, for example, auburn tresses are as red as a tomato or not-red as coal.

By contrast, fuzzy sets admit partial degrees of membership modelled as a function from the universe of discourse to the unit interval[6], termed its *membership function*. The membership function of a fuzzy set X will be denoted μ_X. So for example we might have:

$$\begin{aligned} \mu_{red}(postbox) &= 1 \\ \mu_{red}(snow) &= 0 \\ \mu_{red}(claret) &= 0.7 \end{aligned}$$

3.1.1 Fuzzy logic

In the fuzzy as in the classical case, set theory is intimately related to logic. If fuzzy sets are identified (as above) with the extensions of predicates, it becomes natural to ask how these atomic sentences should be combined with basic sentential connectives $\wedge$ (and), $\vee$ (or) and $\neg$ (not). The classic (cite Zadeh) response is as follows:

$$\begin{aligned} \mu_{true}(A \wedge B) &= min(\mu_{true}(A), \mu_{true}(B)) \\ \mu_{true}(A \vee B) &= max(\mu_{true}(A), \mu_{true}(B)) \\ \mu_{true}(\neg A) &= 1 - \mu_{true}(A) \end{aligned}$$

So, for example, the sentence "The claret is red and so is the postbox" is evaluated has having a truth value of $min(0.7, 1) = 0.7$.

3.1.2 Alphacuts

An equivalent[7] and often convenient representation is to consider the α-cuts of the set in question. The α-cut of a fuzzy set Y at x, $Y_{\downharpoonright x}$, is the set of all elements of the universe of discourse whose membership of Y is greater than or equal to x, i.e.

$$Y_{\downharpoonright x} = \{y \mid \mu_Y(y) \geq x\}$$

Following the example above and assuming a limited universe of discourse, $red_{\downharpoonright 0.7} = \{claret, postbox\}$. α-cuts are especially useful in the context of fuzzy numbers which are be introduced in the following section.

[3] Neither Zadeh's "fuzzy probabilities" nor Jain and Agogino's "Bayesian Fuzzy Probabilities" allow the membership function of a probability term to tail off smoothly to the left. These and other objections are raised and discussed in [11]

[4] Although we hope it *will* be read.

[5] That the world is divided exlusively and exhaustively into two by any predicate is ensured by The Law of the Excluded Middle (nothing is both P and not-P) and The Principle of Bivalence (everything is either P or not-P).

[6] The set of real numbers between 0 and 1.

[7] The α-cut function fully determines the membership function and vice-versa.

q_p	p_l	$\mathrm{P}_f(q_l = none \mid q_p, p_l)$	$\mathrm{P}_f(q_l = few \mid q_p, p_l)$	$\mathrm{P}_f(q_l = many \mid q_p, p_l)$
none	some	1	0	0
	most	1	0	0
	all	1	0	0
few	some	0.5	0.5	0
	most	0.25	0.75	0
	all	0.01	0.99	0
many	some	0.1	0.4	0.5
	most	0.01	0.14	0.85
	all	0	0.1	0.99

Table 6: Classical conditional probabilites of $P(q_l \mid q_p, p_l)$.

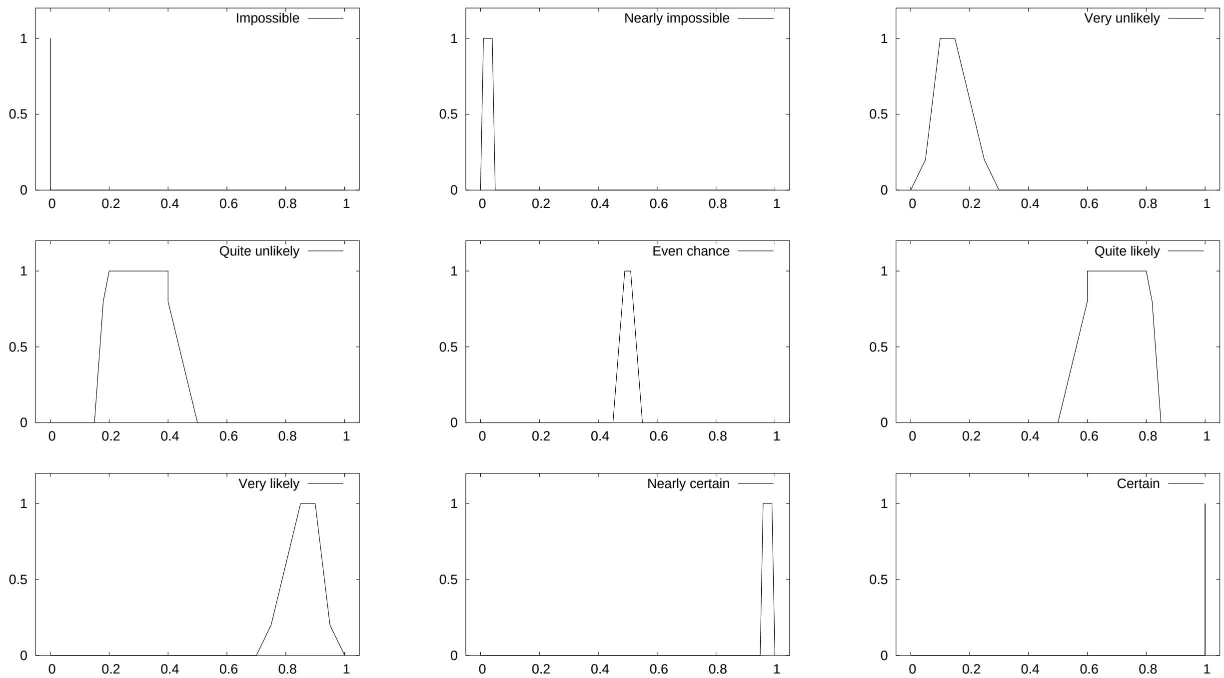

Figure 2: The linguisitic probabilites used in the worked example.

3.2 Fuzzy numbers

A fuzzy numbers are simply fuzzy sets of real numbers with a sensible shape. More formally,

Definition 1 (Fuzzy number) A fuzzy number is a fuzzy set of real numbers, X, whose membership function is:

a) *normal* i.e. $\exists x \in \mathbb{R}$ such that $\mu_X(x) = 1$;
b) *convex* i.e. $\forall x, y, z \in \mathbb{R}$ if $x \leq y \leq z$ then $\mu_X(y) \geq \min(\mu_X(x), \mu_X(z))$; and
c) has a *bounded support* i.e. $\exists N \in \mathbb{R}$ such that $\forall x \in \mathbb{R}$ if $|x| \geq N$ then $\mu_X(x) = 0$.

Note that this also covers "fuzzy intervals". Examples of fuzzy numbers can be found in Figure 2 which defines the linguistic probabilities that will used in the worked example.

3.2.1 Fuzzy arithmetic

The Extension Principle identifies a natural way to extend maps on classical sets to maps on their fuzzy counterparts. The underlying intuition is that a point belongs to the image of a set, A under the extended map, to the extent that its inverse

Definition 2 (Extension Principle) Given a map $f : A \to B$ it's fuzzy counterpart $\tilde{f} : \tilde{A} \to \tilde{B}$ is defined by:

$$\mu_{\tilde{f}(\tilde{a})}(y) = \sup_{x \in A} \{\mu_a(x) \mid f(x) = y\}$$

In other words the possibility of y being in the image of a fuzzy set under an extended funtion is the maximum of the membership values that would have been mapped to it by the original function.

A more palatable explanation? The extension principle may be used to define fuzzy counterparts of standard arithmetic operators. If the standard arithmetic operators are considered as maps from $\mathbb{R}^2 \to \mathbb{R}$

$$\begin{aligned} \mu_{a \oplus b}(z) &= \sup_{x+y=z} \min(\mu_a(x), \mu_b(y)) \\ \mu_{a \otimes b}(z) &= \sup_{xy=z} \min(\mu_a(x), \mu_b(y)) \end{aligned}$$

3.2.2 Ordering fuzzy numbers

It is also possible to use the Extension Principle to induce a partial order on the fuzzy reals. This proceeds from the observation that (in the standard reals) $a \leq b$ if and only if $a = \min(a, b)$. Now define

$$a \preccurlyeq b \text{ iff } \mu_a(z) = \sup_{min(x,y)=z} \min(\mu_a(x), \mu_b(y)) \ \forall z \in \mathbb{R}$$

This can be unpacked as:

$$\forall y \ \exists x \text{ such that } x \leq y \text{ and } \mu_b(y) \leq \mu_a(x)$$

Another partial order is generated by the fuzzy subset relation. Because it is confusing to talk of one fuzzy number being a superset of another we prefer to say that the former *subsumes* the latter. One number subsumes another if is in effect a kind of less precise version of it. Note that since not all fuzzy numbers are comparable (in the sense of the trichotomy law) under either relation, neither constitutes a total order.

A key property of the subsumption relation is that it carries over the arithmetic operators derived using the Extension Principle in the sense of the following Lemma:

Lemma 1 Given an operator $* : \mathbb{R}^n \to \mathbb{R}$ and fuzzy numbers, $a_1, a_2 \ldots a_n, b_1, b_2 \ldots b_n$ such that $a_i \subseteq b_i$ for all $1 \leq i \leq n$ then

$$\circledast(a_1, a_2 \ldots a_n) \subseteq \circledast(b_1, b_2 \ldots b_n)$$

where $\circledast$ is the fuzzy operator derived from $*$ by the Extension Principle.

Proof. By the Extension Principle

$$\begin{aligned} \circledast(a_1, a_2 \ldots a_n)(x) &= \sup_{x=*(x_1,x_2\ldots x_n)} \{ \min_{1 \leq i \leq n} a_i(x_i)\} \\ &\leq \sup_{x=*(x_1,x_2\ldots x_n)} \{ \min_{1 \leq i \leq n} b_i(x_i)\} \\ &= \circledast(b_1, b_2 \ldots b_n)(x) \end{aligned}$$

This result allows complex calculations (such as the Bayesian sum of products expression for joint probability distribution) to be rearranged and computed from partial results.

3.3 Foundations of Probability Theory

The predominant formalisation of probability theory is that provided by Kolmogorov. The standard axioms of probability theory may be found in any introductory text (e.g. [10]) but these are rehearsed below in order to introduce the notation that will be used throughout.

Given an experiment or trial, such as rolling a die, the set of all possible outcomes or *sample space* will be denoted Ω. So, in the die example $\Omega = \{1, 2, 3, 4, 5, 6\}$. Clearly various questions may be asked about the outcome of a trial. Some of these will be elementary, of the form "Was the outcome ω?", but others will be about groups of states. Returning to the die example, one might enquire "Was the outcome an odd number?" That we are not typically interested in individual outcomes, but instead in properties shared by a number of these is captured in the notion of an event space.

Definition 3 (Event space) A set $\mathcal{E}$ is termed an event space on a set Ω of possible outcomes iff

a) $\mathcal{E} \subseteq \mathbb{P}(\Omega)$
b) $\mathcal{E}$ is non-empty.
c) If $A \in \mathcal{E}$ then $A^c \triangleq \Omega \setminus A \in \mathcal{E}$
d) If $A, B \in \mathcal{E}$ then $A \cup B \in \mathcal{E}$

Event spaces are also sometimes called sigma algebras, sigma fields or borel fields.

Note that if $A, B \in \mathcal{E}$ then $A \cap B \in \mathcal{E}$ since $A \cap B = (A^c \cup B^c)^c$, and that this result generalises (by induction) to countable intersections. That an event space is closed under union, intersection and complementation (as implied by conditions c and d) and corresponds to the intuition that if we are interested whether properties P and Q obtain of an outcome, we might also reasonably enquire whether they obtain, conjunctively, disjunctively or not at all. Finally, observe that b entails that both Ω and $\emptyset$ are elements of $\mathcal{E}$ since $(A \cup A^c) = \Omega$.

With the notion of an event space in place it is possible to define the central concept of a probability measure.

Definition 4 (Probability measure) A mapping $\mathrm{P} : \mathcal{E} \to \mathbb{R}$ is termed a probability measure on $(\Omega, \mathcal{E})$ iff

(CP1) $\mathrm{P}(A) \geq 0$ for all $A \in \mathcal{E}$
(CP2) $\mathrm{P}(\Omega) = 1$
(CP3) If A, B are disjoint events in $\mathcal{E}$ (i.e. $A \cap B = \emptyset$) then $\mathrm{P}(A) + \mathrm{P}(B) = \mathrm{P}(A \cup B)$

Where P is a probability measure on $(\Omega, \mathcal{E})$, the tuple $(\Omega, \mathcal{E}, \mathrm{P})$ is termed a probability space.

3.4 Linguistic Probability Theory

Linguistic Probability Theory extends classical probability theory by providing analogues for each of the classical axioms. The advantage of this foundational approach is that the familiar constructs of probability theory such as random variables and distribution functions have natural analogues in the extended theory.

As in the classical case as set of outcomes, Ω, and event algebra, $\mathcal{E}$, are assumed to be given. The concepts of linguistic probability measure measure and linguistic probability space are then defined as follows:

Definition 5 (Linguistic probability measure) A function $\mathrm{P}_f : \mathcal{E} \to \mathcal{F}(\mathbb{R})$ is termed a linguistic probability measure on $(\Omega, \mathcal{E})$ iff

(LP1) $0_\chi \preccurlyeq \mathrm{P}_f(A) \preccurlyeq 1_\chi$ for all $A \in \mathcal{E}$
(LP2) $\mathrm{P}_f(\Omega) = 1_\chi$ and $\mathrm{P}_f(\emptyset) = 0_\chi$
(LP3) If A, B are disjoint events in $\mathcal{E}$ (i.e. $A \cap B = \emptyset$) then $\mathrm{P}_f(A) \oplus \mathrm{P}_f(B) \supseteq \mathrm{P}_f(A \cup B)$

Where P_f is a linguistic probability measure on $(\Omega, \mathcal{E})$, the tuple $(\Omega, \mathcal{E}, \mathrm{P}_f)$ is termed a linguistic probability space.

Like the first two axioms of classical probability theory LP1 and LP2 are essentially arbitrary stipulations. LP1 establishes asserts that, in accordance with intention behind the construction, linguistic probabilities have zero membership outside the chosen quantity space of the unit interval. With the quantity space established, LP2 determines the limit cases, indirectly asserting the completeness of the set of outcomes. Note that whilst the corresponding classical axiom need only specify one of $\mathrm{P}(\emptyset)$ or $\mathrm{P}(\Omega)$ with CP3 determining the other, for reasons discussed below $\mathrm{P}_f(\emptyset)$ and

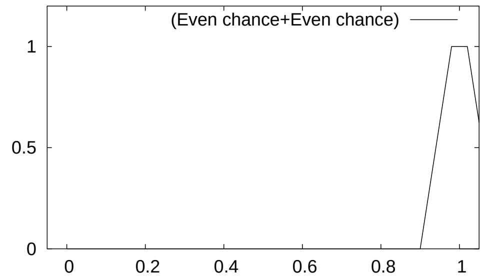

Figure 3: $\mathrm{P}_f(\{Heads\}) \oplus \mathrm{P}_f(\{Tails\})$

$\mathrm{P}_f(\Omega)$ are only loosely coupled by LP3 and must be specified in full.

Whilst the first two axioms are obvious and intuitive,LP3 requires more analysis and justification. The underlying intuition is that computed probabilities are less precise versions of an underlying probability (which, of course, may be fuzzy itself). To see this heuristically, consider the simple case of tossing a coin which is known to be slightly biased. Suppose that, taking the known bias into account, each of the two possible events is assigned a linguistic probability of "even chance" (see Figure 2). Now, $\mathrm{P}_f(\{Heads\}) \oplus \mathrm{P}_f(\{Tails\})$, which is graphed in Figure 3, is not even a linguistic probability! Nevertheless, it is evident that its membership function must dominate that $\mathrm{P}_f(\omega) = 1_\chi$.

Finally, note that Linguistic Probability Theory extends classical probability theory. Where a given linguistic probability measure's range consists solely of embedded real numbers these numbers are in accordance with the classical axioms. A proof of this result and more details about the theory can be found in [11].

4. ILLUSTRATIVE RESULTS

Returning to the This section discusses the results obtained by applying Bayesian Network(s) of section 2 with those obtained by applying the same networks with linguistic probabilities.

4.1 Linguistic network

The linguisitic probabilities used in the specification of the qualitative network are presented in Figure 2, whilst the prior and conditional probabilities specifying the network can be found in Tables LP1, LP2 and LP3.

5. CONCLUSION AND FUTURE WORK

Future work: Translation back into language (cite Javier?) Note about ability to express a total lack of knowledge.

Acknowledgements

This work was partly supported by UK EPSRC award number 00304567. The authors are grateful to Alan Smaill for helpful discussions.

6. REFERENCES

[1] Aitken, C., Taroni, F., and Garbolino, P. A graphical model for the evaluation of cross-transfer evidence in DNA profiles. *Theoretical Population Biology*, 63(3), 2003. To appear.

[2] Balding, D.J. and Donnelly, P. Inference in forensic identification. *Journal of the Royal Statistical Society Series A*, 158:21–53, 1995.

[3] D. V. Budescu and T. S. Wallsten. Consistency in interpretation of probabilistic phrases. *Organizational Behaviour and Human Decision Processes*, 36:391–405, 1985.

[4] Cook, R., Evett, I., Jackson, G., Jones, P., and Lambert, J. A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*, 38:231–239, 1998.

[5] Cook, R., Evett, I., Jackson, G., Jones, P., and Lambert, J. A model for case assessment and interpretation. *Science and Justice*, 38:151–156, 1998.

[6] Cook, R., Evett, I., Jackson, G., Jones, P., and Lambert, J. Case pre-assessment and review in a two-wat transfer case. *Science and Justice*, 39:103–111, 1999.

[7] Dawid, A.P., Mortera, J., Pascali, V.L., and van Boxel, D.W. Probabilistic expert systems for forensic inference from genetic markers. *Scandinavian Journal of Statistics*, 29:577–595, 2002.

[8] Evett, I., Jackson, G., and Lambert, J. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Science and Justice*, 40:3–10, 2000.

[9] Evett, I., Jackson, G., Lambert, J., and McCrossan, S. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice*, 40:233–239, 2000.

[10] G. Grimmet and D. Welsh. *Probability: an introduction*. Oxford University Press, Oxford, 1986.

[11] J. Halliwell and Q. Shen. Towards linguistic probability theory. In *Proceedings of the 11th International Conference on Fuzzy Systems*, pages 596–601, 2002.

[12] P. Jain and A. M. Agogino. Stochastic sensitivity analysis using fuzzy influence diagrams. In R. D. Schachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, pages 79–92. North-Holland, Amsterdam, 1990.

p_s	$\mathrm{P}_f(p_s)$
none	very unlikely
small	quite likely
large	quite unlikely

t_c	$\mathrm{P}_f(t_c)$
none	even chance
some	even chance

p_l	$\mathrm{P}_f(p_l)$
none	almost impossible
few	quite likely
many	quite likely

Table 7: Linguistic prior probabilities $\mathrm{P}_f(p_s), \mathrm{P}_f(t_c)$.

t_c	$\mathrm{P}_f(q_t = none \mid t_c)$	$\mathrm{P}_f(q_t = few \mid t_c)$	$\mathrm{P}_f(q_t = many \mid t_c)$
light	very unlikely	quite likely	very unlikely
medium	very unlikely	even chance	even chance
heavy	nearly impossible	very unlikely	quite likely

Table 8: Linguistic conditional probabilites $\mathrm{P}_f(q_t \mid t_c)$.

q_t	p_s	$\mathrm{P}_f(q_p = none \mid q_t, p_s)$	$\mathrm{P}_f(q_p = few \mid q_t, p_s)$	$\mathrm{P}_f(q_p = many \mid q_t, p_s)$
none	none	certain	impossible	impossible
	small	certain	impossible	impossible
	large	certain	impossible	impossible
few	none	nearly impossible	nearly certain	impossible
	small	very unlikely	very likely	impossible
	large	even chance	even chance	impossible
many	none	impossible	nearly impossible	nearly certain
	small	nearly impossible	very unlikely	very likely
	large	very unlikely	even chance	even chance

Table 9: Linguistic conditional probabilities of $P(q_p \mid q_t, p_s)$.

q_p	p_l	$\mathrm{P}_f(q_l = none \mid q_p, p_l)$	$\mathrm{P}_f(q_l = few \mid q_p, p_l)$	$\mathrm{P}_f(q_l = many \mid q_p, p_l)$
none	some	certain	impossible	impossible
	most	certain	impossible	impossible
	all	certain	impossible	impossible
few	some	even chance	even chance	impossible
	most	quite unlikely	quite likely	impossible
	all	nearly impossible	nearly certain	impossible
many	some	very unlikely	quite unlikely	even chance
	most	very unlikely	very unlikely	very likely
	all	impossible	very unlikely	nearly certain

Table 10: Linguistic conditional probabilites of $P(q_l \mid q_p, p_l)$.

[13] D. Kahneman, P. Slovic, and A. Tversy. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 1985.

[14] National Research Council Governing Board Commitee on the Assessment of Risk. *The handling of risk assessments in NRC reports*. Washington D.C, 1981.

[15] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

[16] T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth. Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115(4):348–365, 1986.

[17] L. Zadeh. Fuzzy probabilities. *Information Processing Management*, 20:363–372, 1984.

[18] A. C. Zimmer. Verbal vs. numerical processing of subjective probabilities. In R. W. Scholz, editor, *Decision making under uncertainty*. North-Holland, Amsterdam, 1983.

[19] A. C. Zimmer. What uncertainty judgements can tell about the underlying subjective probabilities. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*. North-Holland, Amsterdam, 1986.